

CSCI572 HW5

News Sites: Fox News

Name: Zejun Liu

USCID: 6992071555

Steps I followed to complete the assignments:

Step1: Enhancing Your Search Engine with Suggest in Solr

1. Add a search component to solrconfig.xml and tell it to use the SuggestComponent.
2. Add a request handler to solrconfig.xml.
3. Reload the core and change the requesthandler from “/select” to “/suggest”. I typed “califo” in the query and executed it, 5 suggestions have been returned.

Step2: Create “big.txt”

1. Create a maven project in Eclipse to extract text from Fox_News crawled data folder and store the text in a file named “big.txt”.
2. Put this file under the folder /Users/zejunliu/Sites.

Step3: Implement Spelling Correction

1. Download Norvig’s spelling program written in PHP – SpellCorrector.php and put this file under the folder /Users/zejunliu/Sites.
2. Include this file in my hw5.php code which is carried forward from hw4.php.
3. When a user inputs a query, the program will check the big.txt to see if there is a same query/word inside this file. If there is not, the program will consider the most similar word to what you input and display “do you mean:”.

Step4: Implement Auto Complete

1. Auto-Complete feature was added using JQuery, suggestions are displayed in a drop-down fashion as soon as the user begins typing and the suggestions get more streamlined as the user gives in more input, so users can select the corresponding results that is displayed.
2. Excluded stopwords to improve Auto-Complete feature

Step5: Generating Snippet

1. put the simple_dom_html.php file under the folder /Users/zejunliu/Sites for reading the content of html and include this file in the hw5.php code.
2. Generate snippets for websites by matching the words in the words group which is generated by breaking the query with the sentence array which is generated by breaking the content and highlight the query term in the snippet.
3. Add regex to remove special characters.

Analysis of the results:

Spelling Correction Examples

	Misspelt Word	Spelling Correction Word
1	lugage	luggage
2	janary	january
3	compter	computer
4	foood	food
5	califnia	california

Auto Complete Examples

	Character Typed	Suggested Words
1	yell	yell yellow yeltsin yelling yelled
2	unit	unit united until university union
3	din	din dns drink disasters dinosaurs
4	toky	tony toys tokyo toyota toy
5	machi	machi machigan machine machines matching